

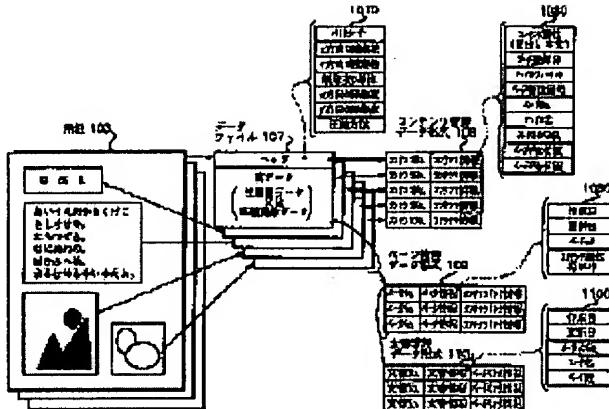
DOCUMENT MANAGING METHOD AND DOCUMENT PROCESSOR FOR REALIZING THE SAME

Patent number: JP8147445
Publication date: 1996-06-07
Inventor: HASHIMOTO HIROHIKO
Applicant: CANON KK
Classification:
- **international:** G06F17/24; G06T1/00; G06F17/24; G06T1/00; (IPC1-7): G06T1/00; G06F17/24
- **european:**
Application number: JP19940281862 19941116
Priority number(s): JP19940281862 19941116

Report a data error here

Abstract of JP8147445

PURPOSE: To provide the document managing method which facilitates the reuse of document image data stored as a digital image and the reconstitution of the data and the document processor for realizing the method. **CONSTITUTION:** This processor is provided with 1st management data 108 for dividing document data 106 corresponding to an attribute showing its contents, regarding data 107 of the respective divided areas as minimum units of document data management, and managing the data in the minimum units, area by area, and the registration and retrieval of the data in the minimum units are managed on the basis of the 1st management data 108. Further, 2nd management data 109 is provided which manages data of plural minimum units, page by page, and document data of page units are further managed on the basis of the link between the 2nd management data 109 and 1st management data 108. Furthermore, 3rd management data 110 is provided which manages data of plural page units by documents, and the document data of document units are managed on the basis of the link between the 3rd management data 110 and 2nd management data 109.



Data supplied from the **esp@cenet** database - Worldwide

【特許請求の範囲】

【請求項1】 デジタル化された文書データを管理する文書管理方法であって、文書データを内容を表す属性に対応して分割して、分割された各領域のデータを文書データ管理の最小単位とし、前記最小単位のデータを各領域毎に管理する第1の管理データを設け、前記第1の管理データに基づいて、前記最小単位のデータの登録及び検索を管理することを特徴とする文書管理方法。

【請求項2】 複数の前記最小単位のデータを各ページ毎に管理する第2の管理データを更に設け、前記第2の管理データと第1の管理データとのリンクに基づいて、更にページ単位の文書データを管理することを特徴とする請求項1記載の文書管理方法。

【請求項3】 複数の前記ページ単位のデータを各文書毎に管理する第3の管理データを更に設け、前記第3の管理データと第2の管理データとのリンクに基づいて、更に文書単位の文書データを管理することを特徴とする請求項2記載の文書管理方法。

【請求項4】 前記第1の管理データと第1の管理データとのリンクを管理する第4の管理データを更に設け、前記第4の管理データに基づいて、更に前記最小単位の文書データを管理することを特徴とする請求項1乃至3のいずれか1つに記載の文書管理方法。

【請求項5】 前記第4の管理データを前記属性単位に設け、前記4の管理データに基づいて、更に同じ属性グループの文書データを管理することを特徴とする請求項4記載の文書管理方法。

【請求項6】 デジタル化された文書データを処理する文書処理装置において、文書データを記憶する第1の記憶手段と、内容を表す属性に対応して分割された各領域単位で文書データを管理する第1の管理データを記憶する第2の記憶手段と、前記第1の管理データに基づいて、前記各領域単位のデータの登録及び検索を管理する第1の管理手段とを備えることを特徴とする文書処理装置。

【請求項7】 前記第2の記憶手段は、複数の前記最小単位のデータを各ページ毎に管理する第2の管理データを更に備え、

前記第2の管理データと第1の管理データとのリンクに基づいて、ページ単位の文書データを管理する第2の管理手段を更に備えることを特徴とする請求項6記載の文書処理装置。

【請求項8】 前記第2の記憶手段は、複数の前記ページ単位のデータを各文書毎に管理する第3の管理データを更に備え、

前記第3の管理データと第2の管理データとのリンクに基づいて、文書単位の文書データを管理する第3の管理手段を更に備えることを特徴とする請求項7記載の文書処理装置。

【請求項9】 前記第2の記憶手段は、前記第1の管理データと第1の管理データとのリンクを管理する第4の管理データを更に備え、

前記第4の管理データに基づいて、前記最小単位の文書データを管理する第4の管理手段を更に備えることを特徴とする請求項6乃至8のいずれか1つに記載の文書処理装置。

【請求項10】 前記第4の管理データを前記属性単位に設け、

前記4の管理データに基づいて、同じ属性グループの文書データを管理する第5の管理手段を更に備えることを特徴とする請求項9記載の文書処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は文書管理方法及びそれを実現する文書処理装置、例えばスキャナ等の画像入力装置で読み取ってデジタル化され、コンピュータで取り扱うことが可能であるようにした画像データ、特に文書画像データを管理する文書管理方法及びそれを実現する文書処理装置に関するものである。

【0002】

【従来の技術】 現在、オフィス情報の大半を占めている紙による情報も、オフィススペースの有効活用と情報の有効活用とのためや、コンピュータ及びスキャナやプリンタ等のコンピュータ周辺装置の普及に伴い、一旦デジタル化して蓄積し、完全な再利用または新しい文書情報の作成のために使用するということが容易に行なわれるようになってきている。

【0003】 このような状況において、オフィスで扱われる紙文書をデジタル情報化することのメリット、即ちスペース効率と容易な情報管理と利用という点を最大限に生かすために、特にデジタル化された紙文書情報をどのように管理し有効活用するかということが大きな問題となっている。かかる状況において、図9に一例を示す従来の文書管理方法では、例えばスキャナ等の画像入力装置で読み取られてデジタル化された文書画像データを管理する。図9で、用紙701は、プリンタやその他の出力装置によって出力された紙を媒体とする文書を有する。この用紙701上の文書は、コンピュータに接続されるスキャナ等の画像入力装置により読み取られデジタルデータに変換されて、ハードディスクや光磁気ディスク等の蓄積装置に保存される。

【0004】 データファイル702は、実際に蓄積装置へ保存される場合のファイル形式の様子を示すものである。通常、デジタルデータとしてコンピュータの機能を使用して保存される場合、コンピュータ上のオペレーテ

イングシステムの機能であるファイルシステムと呼ばれるファイル管理機能のもとに、1つのファイルとして保存されることになる。データファイル702はその時のデータ形式の一例を示したものである。

【0005】データファイル702のヘッダ部7020の情報の一例としては、コンピュータのCPUによって異なる「バイトオーダー」、「x方向の画素数」、「y方向の画素数」、「解像度の単位」、「x方向の解像度」、「y方向の解像度」、「圧縮方法」等の情報がある。実データ部には、「生画像データ」かまたは種々の圧縮方法によって圧縮された「圧縮画像データ」が存在する。データファイル702は、用紙701の1ページに対して1つのデータファイル702を割り当てたものであるが、用紙701の複数ページに対して1つのデータファイル702を割り当てるのももちろん可能である。

【0006】データファイル702からなる文書情報は、ページ管理データと文書管理データとにより管理される。以下に、ページ管理データと文書管理データの形式例を示す。ページ管理データ形式703は、データファイル702の情報において、ページ毎の管理をするためのデータ形式の一例である。ページ管理データ形式703の一例としては、データファイル702のページデータ毎に付けられた「ページNo.」とそれぞれのページのページ情報7030とがある。ページ情報7030は、データが登録された「登録日」、データファイルの保存されているファイル形式を表す「ファイルフォーマット」、文書を管理している番号によってこのページがどの文書に含まれるページであるかを示す「文書No.」、データファイル702の「ファイル名」、データファイル702がファイルシステムのどこに存在するかを示す「ディレクトリパス名」等を含む。また上記に述べたように、用紙701の複数ページに対して1つのデータファイル702を割り当てる場合には、ページ情報として、同一データファイル中の何番目のファイルかを示す「データファイル内No.」が必要になる。

【0007】文書管理データ形式704は、上記に説明したページ管理データ形式703の情報に対して、各々のページのリンク情報を管理することにより、通常は複数のページで構成される一文書ごとの管理を行なうデータ形式の一例である。文書管理データ形式704としては、文書ごとに管理データ内で通番で付けられた「文書No.」と、文書ごとの文書情報7040と、文書を構成しているページの管理を行なうため、文書の何ページ目が上記ページ管理データ形式のどの「ページNo.」に相当するかを示す「ページリンク情報」とがある。文書情報7040は、文書が初めて作成された「作成日」、文書に対して更新がかけられた日を表す「更新日」、文書のバージョン管理を行なう「バージョンNo.」、文書を登録した「ユーザ名」、文書全体の「ペ

ージ数」等を含む。

【0008】図10は、図9に示した管理データによる従来の文書管理方法を実現するためのプログラムの構成を示したものである。文書管理プログラム801は、前述のページ管理データ形式703を有するページ管理データ802と、文書管理データ形式704を有する文書管理データ803を管理して新規作成や更新を行なうと共に、データファイル702の登録、検索、削除等の動作を行なう。

【0009】

【発明が解決しようとする課題】しかしながら、上記従来技術に示した構成では、画像入力装置によってデジタル化された文書画像データの管理単位はあくまでも1ページ単位のデータであり、1ページ内に様々な属性、例えば、「見だし」、「自然画像」、「線図形」などの領域が含まれている場合、これら領域毎の情報やデータを有効に活用することが簡単には出来ない構成となっていた。特に、新聞や雑誌等の場合、1ページ中には通常の「本文文字」の他に「写真」等が数多く含まれるため、

非常に多くの情報が混在し、1ページ単位のデータの管理ではデータの再利用及びデータの再構成という面において、非常に融通性のないデータ管理方法であった。

【0010】本発明は、前記従来の欠点を除去し、デジタル画像として蓄積された文書画像データの再利用及びデータの再構成を容易に行なえるようにする文書管理方法及びそれを実現する文書処理装置を提供すること目的とする。

【0011】

【課題を解決するための手段】上述の課題を解決するために、本発明の文書管理方法は、デジタル化された文書データを管理する文書管理方法であって、文書データを内容を表す属性に対応して分割して、分割された各領域のデータを文書データ管理の最小単位とし、前記最小単位のデータを各領域毎に管理する第1の管理データを設け、前記第1の管理データに基づいて、前記最小単位のデータの登録及び検索を管理することを特徴とする。

【0012】ここで、複数の前記最小単位のデータを各ページ毎に管理する第2の管理データを更に設け、前記第2の管理データと第1の管理データとのリンクに基づいて、更にページ単位の文書データを管理する。また、複数の前記ページ単位のデータを各文書毎に管理する第3の管理データを更に設け、前記第3の管理データと第2の管理データとのリンクに基づいて、更に文書単位の文書データを管理する。また、前記第1の管理データと第1の管理データとのリンクを管理する第4の管理データを更に設け、前記第4の管理データに基づいて、更に前記最小単位の文書データを管理する。また、前記第4の管理データを前記属性単位に設け、前記4の管理データに基づいて、更に同じ属性グループの文書データを管理する。

【0013】又、本発明の文書処理装置は、デジタル化された文書データを処理する文書処理装置において、文書データを記憶する第1の記憶手段と、内容を表す属性に対応して分割された各領域単位で文書データを管理する第1の管理データを記憶する第2の記憶手段と、前記第1の管理データに基づいて、前記各領域単位のデータの登録及び検索を管理する第1の管理手段とを備えることを特徴とする。

【0014】ここで、前記第2の記憶手段は、複数の前記最小単位のデータを各ページ毎に管理する第2の管理データを更に備え、前記第2の管理データと第1の管理データとのリンクに基づいて、ページ単位の文書データを管理する第2の管理手段を更に備える。また、前記第2の記憶手段は、複数の前記ページ単位のデータを各文書毎に管理する第3の管理データを更に備え、前記第3の管理データと第2の管理データとのリンクに基づいて、文書単位の文書データを管理する第3の管理手段を更に備える。また、前記第2の記憶手段は、前記第1の管理データと第1の管理データとのリンクを管理する第4の管理データを更に備え、前記第4の管理データに基づいて、前記最小単位の文書データを管理する第4の管理手段を更に備える。また、前記第4の管理データを前記属性単位に設け、前記4の管理データに基づいて、同じ属性グループの文書データを管理する第5の管理手段を更に備える。

【0015】

【実施例】以下、本発明の実施例を添付図面を用いて詳細に説明する。図1は、本実施例の文書処理装置のシステム構成図である。図中、1は演算・制御用の中央演算部(CPU)、2はROM, RAM等の記憶部であり、例えばROMには文書登録プログラム2aや文書検索プログラム2b等の文書管理プログラムが、RAMにはコンテンツ管理データ2c、ページ管理データ2d、文書管理データ2eが格納されている。3は文書入力あるいはオペレータ指示用のキーを有するキーボード(KBD)、4は文書情報あるいは指示アイコン等を表示する表示部(CRT)、5は表示された指示アイコンをクリックしたりカーソルを移動させるマウス、6はRAMにロードされて実行されるプログラム等を格納するフロッピーディスク(FD)、7はRAMの各管理データに基づいて文書管理プログラムで管理される文書のデータファイル7aを蓄積するハードディスク(HDD)や光磁気ディスク(MO)等の外部記憶装置、8は用紙から文書を読み取るスキャナ、9は文書情報の送受信をする通信インターフェイス、11は上記各部位を接続するバスである。

【0016】図2の用紙101上の文書(文書101と呼ぶこととする)は、文書1ページ内にどの様な属性をもつ領域があるかを示した例である。文書101には、「見だし」102、「本文」103、「自然画像」10

4、「線図形」105の属性をもつ領域が存在している。これら領域の属性は文書を管理するユーザがユーザ毎の使用形態に基づいて設定し、また領域を指定するものでもよいし、プログラムで自動的に属性判別、領域判別ができるものでもよい。

【0017】<実施例1>図3は、図2に示す用紙のデジタル文書画像データからなる文書を管理するための実施例1の管理データの構成を示すものである。図3の用紙106は、デジタル画像データに変換される紙を媒体とした文書であり、図9で説明した用紙701と同様のものである。用紙106上の文書(文書106と呼ぶこととする)は、コンピュータに接続されるスキャナ8等の画像入力装置によりデジタルデータに変換され、ユーザかまたはプログラムにより図2の文書101に示したような属性判別情報(「見出し」等)と領域判別情報とが得られる。

【0018】図3に示すデータファイル107は、前記それぞれの属性の領域ごとにデータとして作成されたファイルであり、実際にハードディスクや光磁気ディスク記憶装置7等のコンピュータに接続された蓄積装置に保存される場合の、ファイル形式の様子を示した一例である。データファイル107のヘッダ部1070の情報の一例としては、コンピュータのCPUによって異なる「バイトオーダ」、「x方向の画素数」、「y方向の画素数」、「解像度の単位」、「x方向の解像度」、「y方向の解像度」、「圧縮方法」等を含む。実データ部には、「生画像データ」かまたは種々の圧縮方法によって圧縮された「圧縮画像データ」が存在する。データファイル107は用紙106の1ページの中の1コンテンツ30に対して1つのデータファイル107を割り当てたものであるが、用紙1ページに対して1データファイル107を割り当てることももちろん可能である。ただし、この場合もデータ管理の最小単位は、上記属性の領域毎に作成されたデータであることには変わりない。

【0019】コンテンツ管理データ形式108は、データファイル107の情報において、コンテンツ毎の情報を管理するためのデータの形式の一例である。コンテンツ管理データ形式108としては、データファイル107のコンテンツデータ毎に付けられた「コンテンツNo.」とそれぞれのコンテンツ情報1080とを有している。

【0020】コンテンツ情報1080は、「見だし」や「本文」といったコンテンツの属性を示す「コンテンツ属性」、データが登録された「登録日」、データファイルのファイル形式を示す「ファイルフォーマット」、ページを管理している番号によって、このコンテンツがどのページに含まれるコンテンツであるかを示す「ページNo.」、データファイル107の「ファイル名」、データファイル107がファイルシステムのどこに存在するかを示す「ディレクトリパス名」、データファイル1

07のデータが、ページのどの位置に存在していたかを示す座標「ページ内x位置」、「ページ内y位置」等を含む。

【0021】ページ管理データ形式109は、上記に説明したコンテンツ管理データ形式108の情報に対して、各ページに含まれる各々のコンテンツのリンク情報を管理することにより、通常は複数のコンテンツによって構成される1ページごとの管理を行なうデータベースの形式の一例である。ページ管理データ形式109としては、ページごとに管理データ内でつけられた「ページNo.」と、ページごとのページ情報1090と、ページを構成しているコンテンツの管理を行なうため、上記コンテンツ管理データ形式108に示したどの「コンテンツNo.」がページ内に含まれるかを示した情報である「コンテンツリンク情報」とを有する。

【0022】ページ情報1090は、ページが作成された「作成日」、更新された「更新日」、ページの「バージョン」、ページ内に含まれるコンテンツの属性毎にいくつコンテンツが含まれるかを示した「コンテンツ属性毎カウント」等を含む。文書管理データ形式110は、上記に説明したページ管理データ形式109の情報に対して、各々のページのリンク情報を管理することにより、通常は複数のページで構成される1文書ごとの管理を行なうデータベースの形式の一例である。文書管理データ形式110としては、文書ごとに管理データ内で通常で付けられた「文書No.」と、文書ごとの文書情報1100と、文書を構成しているページの管理を行なうため、文書の何ページ目が上記ページ管理データ形式のどの「ページNo.」に相当するかを示す「ページリンク情報」とがある。

【0023】文書情報1100は、文書が初めて作成された「作成日」、文書に対して更新がかけられた日を表す「更新日」、文書のバージョン管理を行なう「バージョンNo.」、文書を登録した「ユーザ名」、文書全体の「ページ数」等を含む。図4は、図3に示した各管理データを基に本実施例の文書管理方法を実現するためのプログラムの構成図を示したものである。

【0024】文書管理プログラム201は、コンテンツ管理データ形式108を有するコンテンツ管理データ202、ページ管理データ形式109を有するページ管理データ203、文書管理データ形式110を有する文書管理データ204を管理しながら、文書のデータファイルの追加、削除、更新等の管理を行なう。図5は、図4に示した文書管理プログラム201が文書の登録を行なう場合の本実施例のフローチャートを示したものである。なお、用紙の文書を画像入力装置で読み取ってデジタルデータに変換された文書画像データが既に用意されているものとする。

【0025】ステップS301で1文書の新規登録がスタートし、次のステップS302で、1ページ文の入力

イメージデータが文書管理プログラム201に渡される。ここで、入力されたデータに対して、ユーザがユーザ毎の使用形態に基づいて、属性を設定し属性毎の領域を指定するか、またはプログラムで自動的に属性を設定して属性毎の領域を指定させてもよい。

【0026】次にステップS303では、分割された領域に基づいて領域毎のデータファイル107を作成し、ステップS304では、データファイル107の管理情報をコンテンツ管理データ202に登録する。コンテンツ管理データ202に基づき、ステップS305でページ管理データ203への登録を行なう。ステップS306で次のページがある場合はステップ302に戻って、ステップ302からステップ305を繰り返す。次のページがない場合は、ステップS307で、ページ管理データ203に基づき文書管理データ204に登録して、ステップS308の文書登録の終了となる。

【0027】図6は、図4に示した文書管理プログラム201が文書検索を行なう場合の本実施例のフローチャートである。例えば、文書内のコンテンツの属性として、文書101のように自然画像104が指定されている場合、以下図6を用いて説明するように文書検索が可能となる。ステップS401より自然画像をもつ文書を検索スタートすると、ステップS402で、図4に示す文書管理プログラム201がコンテンツ管理データ203の検索を「自然画像」の属性で行ない、ステップS403で検索結果をユーザに提供し、ステップS404で検索を終了する。ここで、ユーザに提示する検索結果としては、自然画像が含まれ文書管理データ204で管理されている文書の情報、ページ管理データ203で管理されているページの情報、コンテンツ管理データ202で管理されているコンテンツの情報がある。

【0028】<実施例2>図7は、本発明による他の文書管理方法を実現する管理データ例と、検索表示例を示すものである。図7の(a)は、コンテンツ管理データ202に基づき、コンテンツの1つの属性の例である「見だし」用のデータを再構成した場合の見だし管理データ形式501の一例を示したものである。見だし管理データ形式501は、上記に示したコンテンツ管理データ形式108の1つの項目である「コンテンツ属性」の項目から、「見だし」の属性をもつものを抜き出して並べたものを「見だし」専用の管理データ形式としたものである。

【0029】図7の(b)に示す502は、見だし管理データ形式501によって作成されたデータをもとにして検索された見だしデータの表示例である。各属性専用の管理データを構築することにより、検索が高速に行なわれるようになる。

<実施例3>図8は、本発明による更に他の文書管理方法を実現する管理データ例と、コンテンツのリンク例を示すものである。

9

【0030】図8の(a)は、コンテンツとコンテンツとのリンク情報を新たに作成した場合のコンテンツリンク管理形式の1例である。コンテンツリンク管理形式601は、文書管理プログラム201でコンテンツのリンク情報を管理するために付けられた「リンクNo.」と、そのリンクの内容を示す2つの「コンテンツNo.」からなる。

【0031】図8の(b)では、コンテンツ602に対して、コンテンツ603とコンテンツ604とがリンクされている様子を示す。ユーザは、ページ単位毎を基準とするのではなく、そのページに含まれるコンテンツ毎のリンク情報を設定することができる。コンテンツ602からコンテンツ603, 604を引き出すことができるし、逆にコンテンツ603, 604からコンテンツ602を容易に引き出すことができる。

【0032】尚、管理対象データの元となるデジタル画像データは、スキャナ等の画像入力装置によって読み込まれたものだけでなく、コンピュータ上のワードプロセッサソフトウェアやディスクトップパブリッシング(DTP)ソフトウェアなどによって作成されたソフトウェアデータを画像データに変換したものでもよい。又、本発明は、複数の機器から構成されるシステムに適用しても1つの機器から成る装置に適用しても良い。また、本発明は、システム或は装置にプログラムを供給することによって達成される場合にも適用できることはいうまでもない。

【0033】

【発明の効果】以上説明したように、本発明にかかる文書管理方法及びそれを実現する文書処理装置によれば、データの再構成や再利用に対して柔軟に対応できる上、ユーザにとっては、ページ毎の管理を意識することなく、データとして何があるのかという情報だけを意識すればよいため、データ再利用の可能性が高まり、データの価値を増す効果がある。

【図面の簡単な説明】

【図1】本実施例の文書処理装置のシステム構成図である。

【図2】本実施例の文書管理方法を説明するための用紙(文書)の例である。

【図3】実施例1による文書管理方法を実現する管理データを説明するための構成図である。

【図4】実施例1による文書管理方法を示すプログラムの構成図である。

【図5】実施例1による文書登録方法の手順を示すフローチャートである。

【図6】実施例1による文書検索方法の手順を示すフローチャートである。

【図7】実施例2による文書管理方法を実現する管理データ及び検索表示例を示す図である。

【図8】実施例3による文書管理方法を実現する管理データ及びコンテンツのリンク例を示す図である。

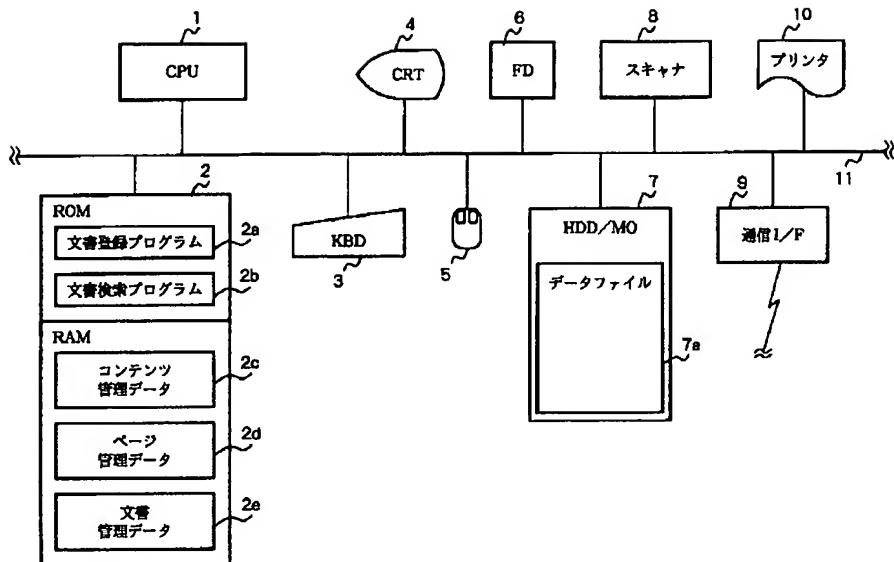
【図9】従来の文書管理方法を実現する管理データの構成例を示す図である。

【図10】従来の文書管理方法を示すプログラムの構成図である。

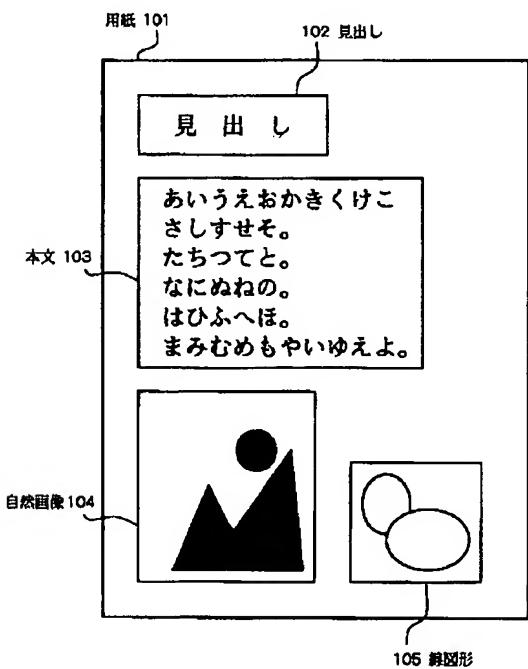
【符号の説明】

- 1 中央演算装置(CPU)
- 2 記憶部(ROM, RAM)
- 2a 文書登録プログラム
- 2b 文書検索プログラム
- 2c コンテンツ管理データ
- 2d ページ管理データ
- 2e 文書管理データ
- 3 キーボード(KBD)
- 4 表示部(CRT)
- 5 マウス
- 6 フロッピーディスク(FD)
- 7 外部記憶装置(HDD, MO)
- 7a データファイル
- 8 スキャナ
- 9 通信インターフェイス
- 10 パス
- 101, 106 用紙(文書)
- 102 見出し
- 103 本文
- 104, 602 自然画像
- 105 線図形
- 107 データファイル
- 108 コンテンツ管理データ形式
- 109 ページ管理データ形式
- 110 文書管理データ形式
- 201 文書管理プログラム
- 202 コンテンツ管理データ
- 203 ページ管理データ
- 204 文書管理データ
- 501 見出し管理データ形式
- 502 検索された見出しデータの例
- 601 コンテンツ管理データ形式
- 603, 604 コンテンツ

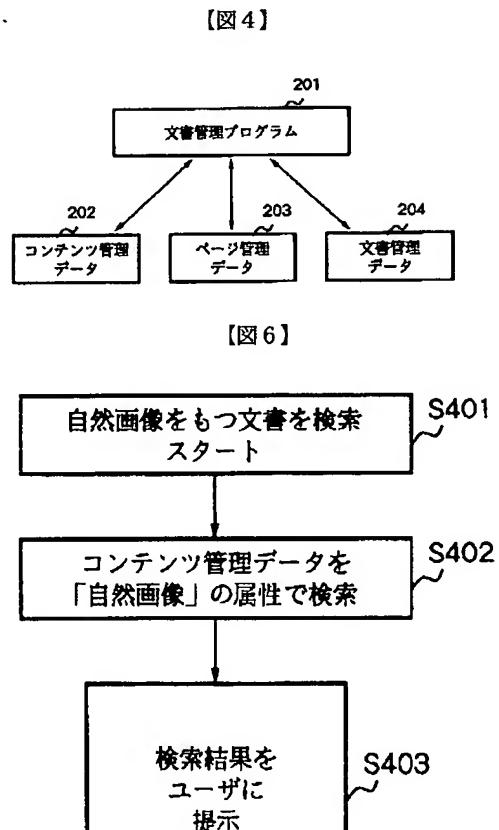
【図1】



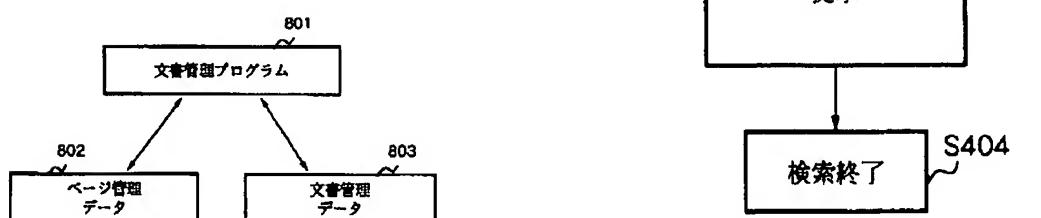
【図2】



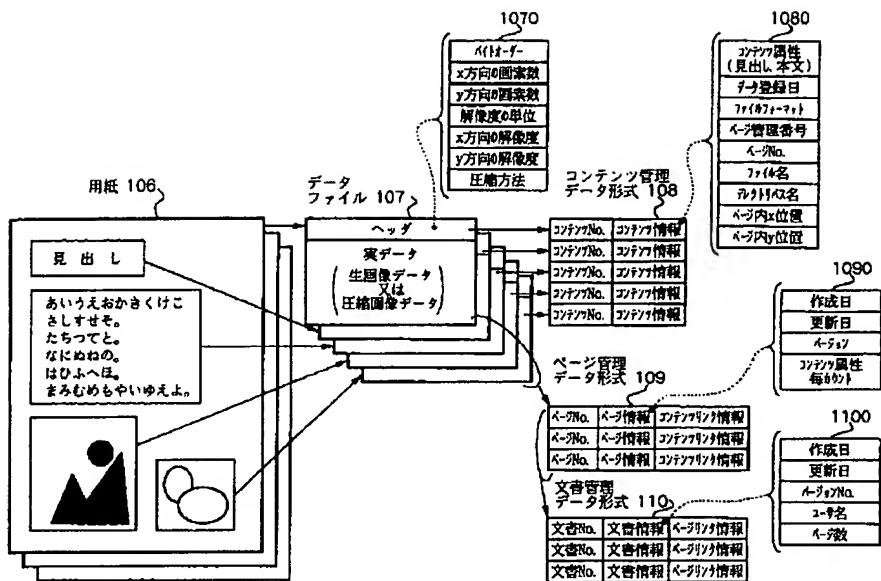
【図10】



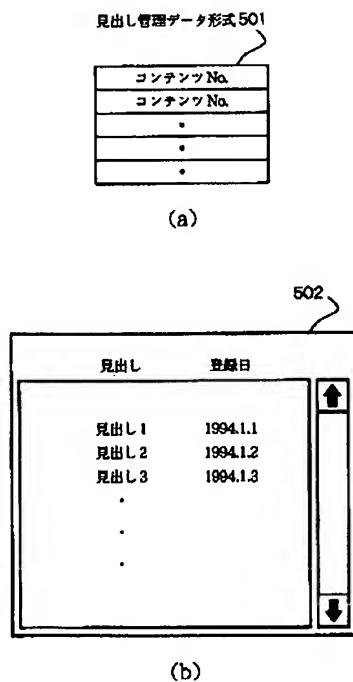
【図6】



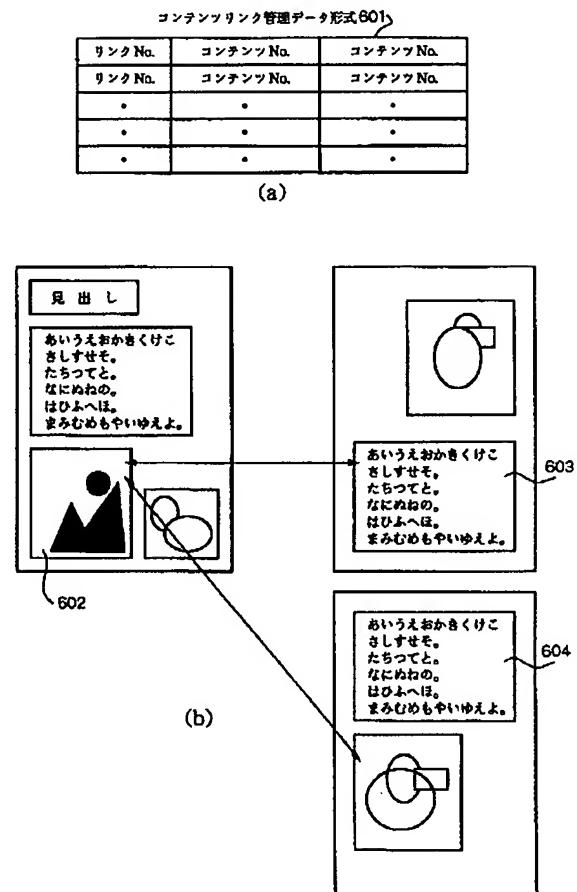
【図3】



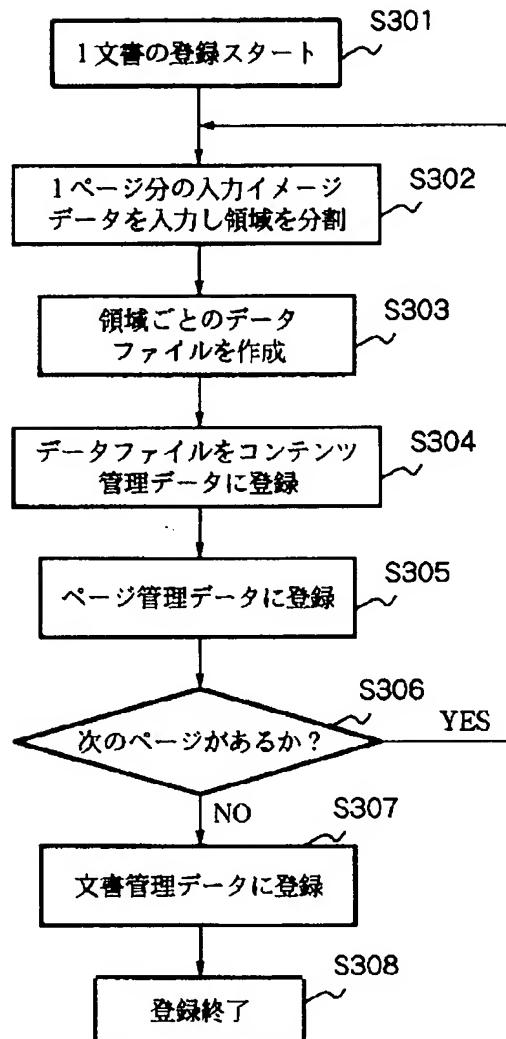
【図7】



【図8】



【図5】



【図9】

